ANALYSIS AND CLASSIFICATION OF LEAF DISEASE BASED ON LEAF TEXTURE USING SUPPORT VECTOR MACHINE

Anggita Puspawardani
Dr. Budi Utami Fahnun, SKom., MM

anggitpuspa07@gmail.com
bufahnun@gmailcom
Department of Informatics, Faculty of Industrial Technology, Gunadarma University

ABSTRACT

Leaf diseases such as Alternaria Alternata, Anthranose, Bacterial Blight, and Cercospora Leaf Spot can be detected using digital image processing applications. One of the steps taken in digital image processing is texture classification which is an important step in digital image processing by identifying an image that provides information about an extraction's value of the texture feature. The Gray Level Co-occurrence Matrix (GLCM) method is one of the methods that produces value of texture features such as Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis and Skewness. From all the extraction values of this texture feature can be classified using the Support Vector Machine (SVM) method so as to produce the name of the type of leaf disease. SVM has a kernel which can provide accurate values. From the research that has been done, it can be proven by the classification of leaf diseases which get an average accuracy rate of 98,349%.
Keyword: Leaf Diseases, Support Vector Machine, Classification, Texture.

INTRODUCTION

According to (Coley-Smith, 2008) plant diseases are defined as abnormal growth. Disease is the result of several disorders in the life process of plants. The cause of plant diseases can come from biotic and abiotic diseases. Biotic diseases are caused by living organisms, such as fungi, bacteria, and viruses. Plant diseases can be identified on the leaves that have different colors, and texture of the leaves themselves. Leaf disease is usually characterized by patches on the leaf surface. Characteristics such as dark brown and red around the spots. From the characteristics of leaf diseases can be classified by type of leaf disease. Examples of leaf diseases such as Alternaria Alternata, Anthranose, Bacterial Blight, and Cercospora Leaf Spot.

Research by (Muchtar and Cahyani, 2018) conducted research on the classification of leaves using image processing techniques on the grounds that it can facilitate the classification. Leaves are part of a plant that can be classified based on texture. The texture feature can work on leaves that are intentionally damaged or their size is too large which sometimes makes the process of image acquisition more difficult. The method used in this study is a combination of the Gabor Filter and Co-occurrence Matrices methods to produce the most appropriate texture features for classifying leaves. The classification uses SVM with a 5-fold cross validation system which shows that the goal of the Gabor Co-occurrence method can achieve an average accuracy of up to 89.83%.

Research by (Vidyashanakara and Kumar, 2018) classifies leaves using the Gray Level Co-occurrence Matrix (GLCM) to extract features based on the texture of the leaves and the Support Vector Machine (SVM) are used to improve the accuracy of the leaf classification. Creating a database for leaves is very useful in speed and efficiency in classification and recognition which are important steps for conservation.

Research by (Sollapure, Karadiguddi, Hanasi, Daddi and Kale, 2018) explains that leaf disease can affect productivity, quality, and efficiency, so that it can affect the growth of plants. To overcome the disease of the leaves must be done at an early stage, then MATLAB is used as an image processing process. The method used is SVM which has a linear kernel that is used to identify diseases. The algorithm used is the Multi Support Vector Machine that can identify many diseases by increasing plant efficiency and productivity.

RESEARCH METHODS

In this research method, the flow process of classification of leaf disease research is shown in Figure 1.

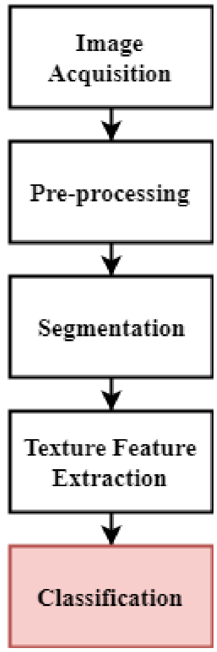


Figure 1 Process Flow Classification of Leaves Disease

Below is an explanation of each path, of the disease classification process in the leaves, as follows:

Image Acquisition

The image acquisition process is the initial stage to take or obtain digital images using a certain additional device or tool. The image used in this study is a leaf image taken from a MATLAB community member named Manu BN in JPG format. The original image used has patches on the surface of the leaf, which indicates the leaf is diseased. Figure 2 is a display of the diseased leaf image.

Figure 2 Acquisition of Diseased Leaf Image.

Pre-Processing

Pre-processing is used to enhance image data by removing the background, this feature enhances the image for processing and analysis. Enhance contrast process is done to get a new RGB value with better contrast. With good contrast in the image, can increase the sharpness of the color of objects in the image, objects that are clearly visible in the image can help the process of image segmentation (Dwi, 2017). The enhance contrast result will then enter the segmentation process. General commands on Matlab R2018a for enhance contrast, are as follows:

I4 = imadjust(I3,stretchlim(I3));

where, stretchlim is the process of stretching RGB values based on the maximum and minimum values of the RGB value. Figure 3 shows the pre-processing results of Bacterial Blight.

(a) (b)

Figure 3 Pre-processing results from Bacterial Blight Image, (a) original image, (b) enhance contrast

Segmentation

Segmentation is used to divide a region into several parts with the same characteristics or have some similarities so that it is easily analyzed. Segmentation is done to classify the leaves that are affected and those that are not affected. The K-Means clustering method is used to partition images into clusters, where at least one part of the cluster contains an image with the main area of the affected part (Min and Htun, 2018). An example of K-Means clustering on leaves infected with Alternaria Alternata can be seen in Figure 4.

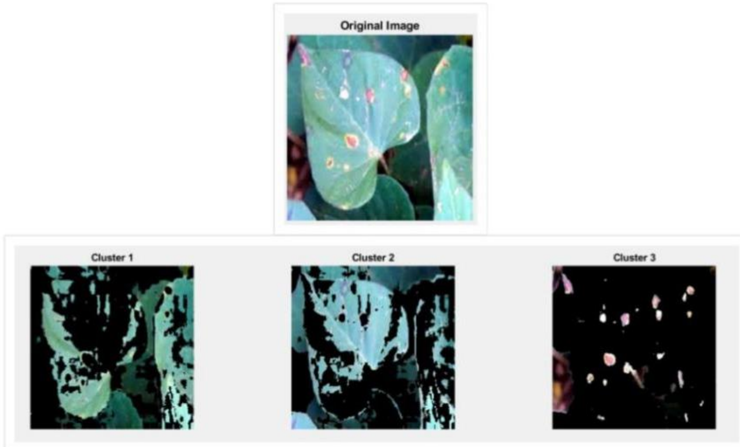


Figure 4 Segmentation of Diseased Leaf Image.

Feature Extraction

After segmentation of the infected area, various features are extracted to provide information about the infected area. The texture feature is used to describe the region. Texture is one of the most important features that can be used to classify and recognize objects (Haralick, 2002).

Gray Level Co-occurrence Matrix (GLCM) is a feature extraction method used for textures that produces statistical calculations. Texture features can be calculated from the resulting GLCM, such as contrast, correlation, energy, entropy, homogeneity, mean, standard deviation, RMS, variance, smoothness, kurtosis, skewness, and IDM.

Classification

After extracting the texture features, the classification is done using the Support Vector Machine (SVM). SVM is a supervised learning that supports hyperplane in high dimensional space with an algorithm that can analyze data and recognize patterns, used for classification and regression analysis (Min and Htun, 2018).

The training and validation process is an important step in developing using SVM. The dataset for the training and validation process consists of two parts, namely the training feature used to train the SVM model and the test feature used to verify the accuracy of the trained SVM model. Then, the results of the type of disease with accuracy and percentage of the area affected by the disease are evaluated by the ratio of leaf disease data (Min and Htun, 2018).

Gray Level Co-occurrence Matrix

In statistical texture analysis, feature extraction is calculated from the statistical distribution of combinations of intensities observed at positions determined by each other in the image. According to the number of intensity points (pixels) in each combination, statistics are classified into the first level, second level, and higher statistics (Albregtsen, 2008).

One of the most popular statistical methods for measuring image information based on texture is the Gray Level Co-occurrence Matrix (GLCM). The GLCM method provides proportional texture information from an image that can only be obtained from two pixels. The Gray Level Co-occurrence Matrix is a method proposed by Robert Haralick that attempts to describe the texture statistically from a sample of how certain gray levels occur in relations between gray levels. (Rakhmawati, Pranoto and Setyati, 2018).

GLCM has 14 features proposed by Haralick. But not all features are used. Generally, only 5 features are used in various studies, namely energy, entropy, contrast, correlation and homogeneity. This is because these five features are the main features in GLCM. The rest are derived from these 5 features (Sahaduta and Lubis, 2013). The five features of GLCM used in this study are as follows:

1. *Entropy*:

$$\boldsymbol{f_1} = -\sum_{i}\sum_{j} \boldsymbol{p(i,j)\log(p(i,j))} \qquad (\mathbf{1})$$

2. *Energy*:

$$\boldsymbol{f_2} = \sum_{i,j} \{\boldsymbol{p(i,j)}\}^2 \qquad (\mathbf{2})$$

3. *Contrast*:

$$f_3 = \sum_{i,j} |i-j|^2 p(i,j) \quad (3)$$

4. *Correlation*:

$$f_4 = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (4)$$

5. *Homogeneity*:

$$f_5 = \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \quad (5)$$

6. *Mean*:

$$f_6 = \sum_{i=0}^{G-1} ip(i) \quad (6)$$

7. *Standard Deviation*:

$$f_7 = \sqrt{\sigma_i^2} \quad (7)$$

8. RMS:

$$f_8 = \sqrt{\frac{1}{N} \sum_{n=1}^{N} |X_n|^2} \quad (8)$$

9. *Variance*:

$$f_9 = \sum_i \sum_j (i - \mu)^2 p(i,j) \quad (9)$$

10. *Smoothness*:

$$f_{10} = 1 - \frac{1}{1+\sigma^2} \quad (10)$$

11. *Skewness*:

$$f_{11} = \sum_{i=0}^{N-1} (i-\mu)^3 P(i) \quad (11)$$

12. *Kurtosis:*

$$f_{12} = \sum_{i=0}^{N-1} (i - \mu)^4 P(i) \quad (12)$$

13. *Inverse Difference Moment*:

$$f_{13} = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j) \qquad (13)$$

Support Vector Machine

Support vector machine (SVM) is a machine learning method for pattern recognition. The first SVM algorithm itself was invented by Vladimir Vapnik. SVM is included in supervised learning which can be used for classification and analysis problems. SVM works on statistical learning and can produce strong, accurate, and effective results with a smaller number of training data samples. SVM is a machine learning that conducts training using training datasets and generalizes and makes predictions from new data. SVM classifies data into two different classes by making decision boundaries are commonly called hyperplane (Rakhmawati, Pranoto and Setyati, 2018).

The main principle of using SVM is to find the best hyperplane that functions as a separator of two classes in the input space. The hyperplane can be a line in two dimensions and can be a flat plane in multiple planes. Like the example, in Figure 5 where the input space has two different objects, such as a circle and positive. SVM defines a boundary between the two objects in which circles and triangles are separated by lines called hyperplane optimised so that the object can maximize margins between two classes of objects (Stipaničev, 1994).

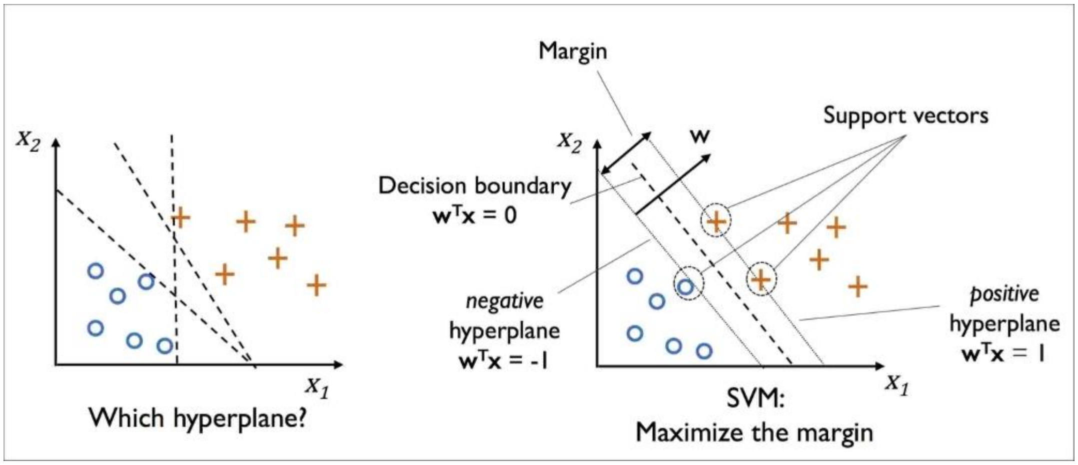


Figure 5 Classification of Support Vector Machines (SVM) (Mirjalili and Raschka, 2017)

Support Vector Machine (SVM) uses a linear model as a decision boundary with the following general form:

$$y(x) = w^T \emptyset(x) + b \tag{14}$$

where x is the input vector, w is the weight parameter, $\emptyset$ (x) is the basis function, and b is a bias.

Besides classifying with two classes, Support Vector Machine has a function which can classify those who have more than 2 classes called Multi-class classification. While the standard SVM is designed for two-class classification problems. There are two types of SVM that deal with multi-class problems, namely one-against-one and one-against-all. One-against-one where SVM binaries are trained for every two data classes to build decision functions. Then k (k-1) / 2 is a decision function for k-class problems. Whereas one-against-all where there is an SVM binary from each class to isolate members from one class from another class (Kulinavar and Hadimani, 2017).

DISCUSSION

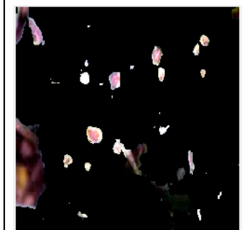
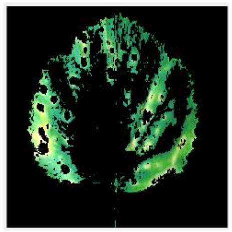
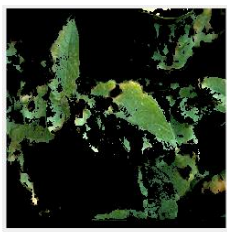
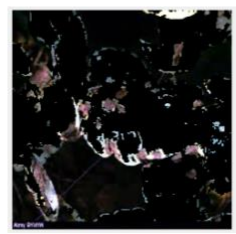
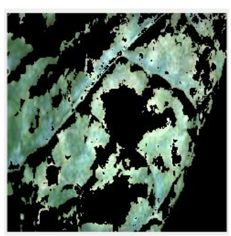
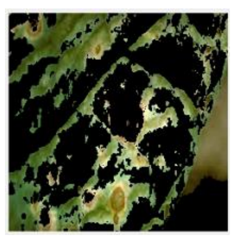
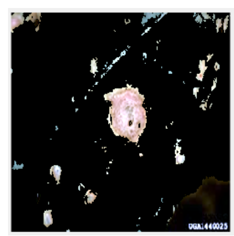
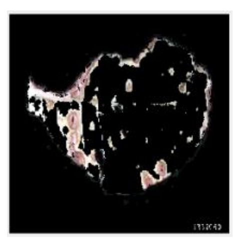
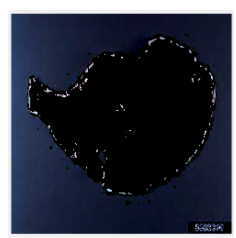
The results of trials from research conducted produce images of the results of the classification of leaf diseases along with the level of accuracy of the detection in the image. The test material is the leaf image. There are four types of leaf image diseases tested, namely Alternaria Alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot.

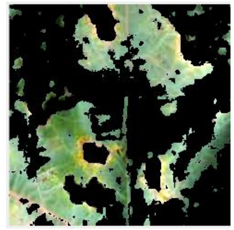
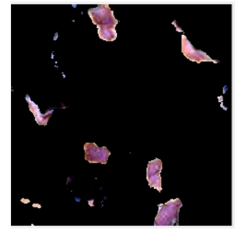
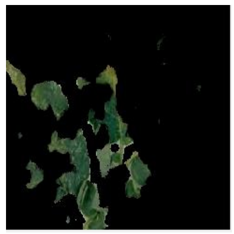
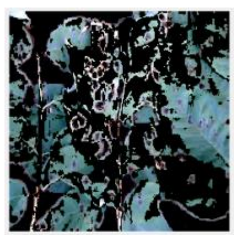
Image Segmentation Results

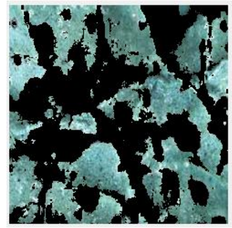
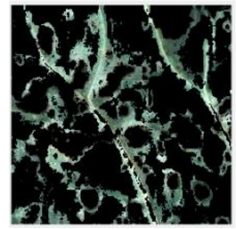
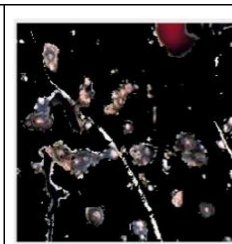
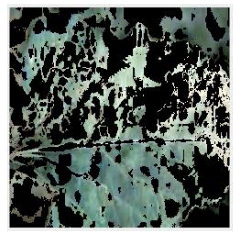
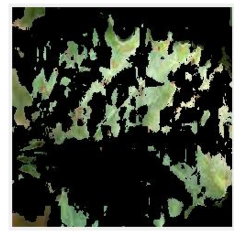
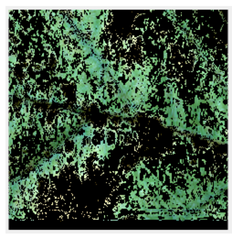
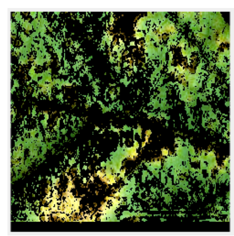
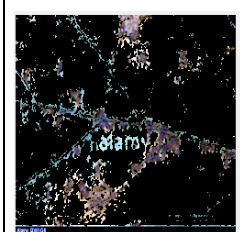
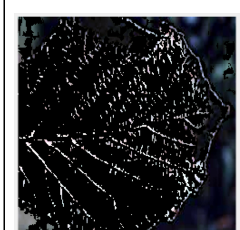
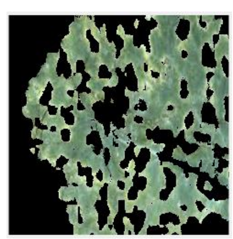
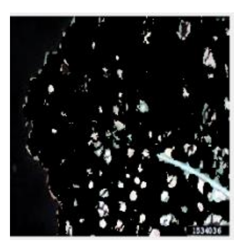
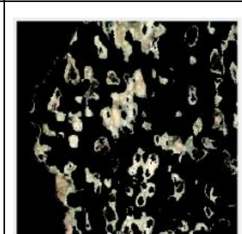
Based on the results of segmentation that has been done on Alternaria Alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot images can be seen in Table 1.

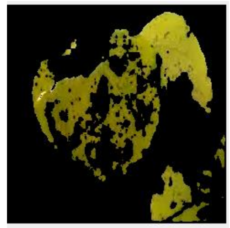
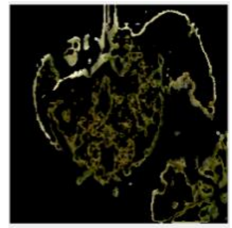
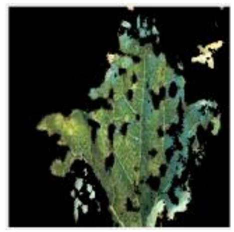
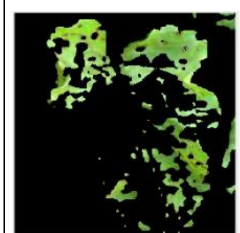
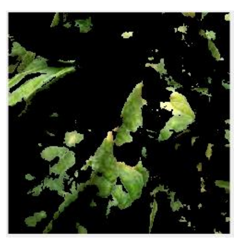
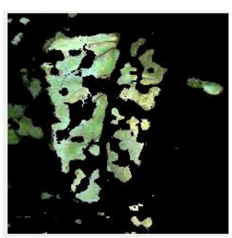
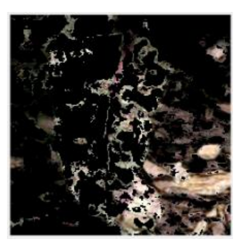
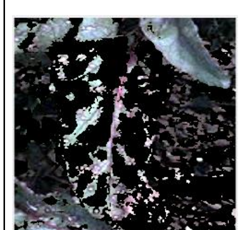
The four types of imagery that have been detected by the leaf surface area with the K-Means Clustering algorithm produce three different clusters. These three clusters are one of the images with the main area affected by the disease.

Table 1 Results of Diseased Leaf Image Segmentation

Image	Original Image	Cluster 1	Cluster 2	Cluster 3
1				
2				
3				
4				
5				

6				
7				
8				
9				
10				

11				
12				
13				
14				
15				

16				
17				
18				
19				
20				

Extraction Feature Results

Seen in table 2 is the value of feature extraction on each image that has passed the segmentation process. This stage is the result of one of the 20 clusters of images that have passed the segmentation stage, the selected cluster is the cluster that has been detected as a leaf spot disease.

Table 2 Results of Extraction of Diseased Leaf Image Features

Image	A	B	C	D	E	F	G	H	I	J	K	L	M
1	10.60	41.47	1.19	3.82	1594.45	1	22.3	4.42	255	0.41	0.84	0.84	0.97
2	17.13	35.54	2.84	10.45	1163.81	1	27.54	4.67	255	0.51	0.71	0.89	0.97
3	18.73	48.36	3.36	7.20	2261.74	1	13.17	3.21	255	0.86	0.76	0.69	0.93
4	16.44	55.68	1.30	4.32	2845.02	1	12.29	3.29	255	0.89	0.82	0.81	0.96
5	16.90	55.22	1.31	4.64	2827.49	1	12.12	3.24	255	1.05	0.78	0.79	0.95
6	78.45	95.51	5.45	11.64	6492.19	1	1.75	0.72	255	0.50	0.96	0.28	0.92
7	125.29	113.64	4.55	11.72	8577.4	1	1.08	-0.13	255	0.63	0.97	0.31	0.95
8	13.00	38.78	1.44	5.06	1340.23	1	12.64	3.17	255	0.34	0.74	0.75	0.96
9	26.90	60.83	2.47	6.94	3467.27	1	6.64	2.21	255	1.04	0.81	0.61	0.92
10	28.91	61.26	2.65	7.51	3570.52	1	6.64	2.15	255	1.38	0.77	0.56	0.89
11	32.08	66.62	2.29	7.14	4219.42	1	5.16	1.87	255	1.48	0.78	0.58	0.92
12	22.76	57.33	1.98	6.21	3126.7	1	8.14	2.50	255	2.18	0.58	0.63	0.90
13	20.58	53.23	2.83	7.25	2771.5	1	11.88	3.10	255	2.64	0.42	0.67	0.91
14	9.74	38.48	0.88	3.51	1388.83	1	19.34	4.10	175	0.27	0.87	0.85	0.97
15	33.70	70.13	2.21	6.65	4382.64	1	4.31	1.73	255	1.97	0.76	0.58	0.89

16	71.91	83.07	5.12	11.46	5683.72	1	1.82	0.64	255	0.48	0.95	0.26	0.94
17	39.89	65.61	3.42	8.79	3942.17	1	3.78	1.43	255	1.20	0.83	0.41	0.89
18	51.69	70.93	4.98	11.12	4409.03	1	3.71	1.37	255	0.83	0.89	0.30	0.91
19	42.05	75.78	3.18	8.45	5343.01	1	3.65	1.51	255	1.30	0.86	0.50	0.92
20	50.79	71.60	4.68	10.51	4813.8	1	3.13	1.19	255	1.32	0.83	0.33	0.88

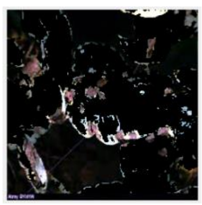
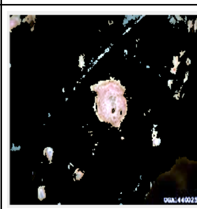
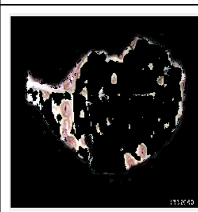
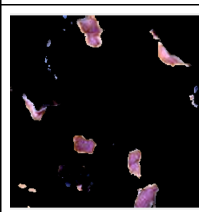
Information:

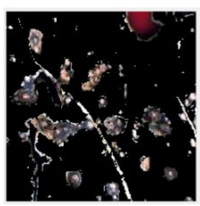
A : Mean
B : Standard Deviation
C : Entropy
D : RMS
E : Variance
F : Smoothness
G : Kurtosis
H : Skewness
I : IDM
J : Contrast
K : Correlation
L : Energy
M : Homogeneity

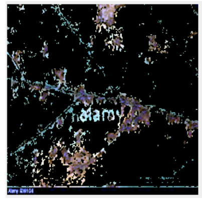
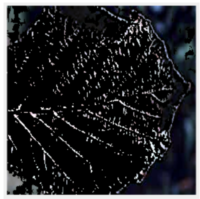
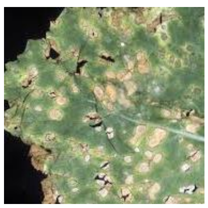
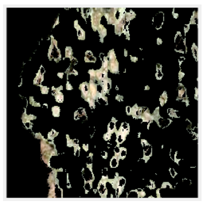
Image Classification Results

Then for the results of the classification and accuracy of the image above can be seen in table 3. The results of this classification are images that have been extracted from previous features, from the values of extraction this feature provides information about the type of leaf disease. While the results of the level of accuracy states the performance of the Support Vector Machine method.

Table 3 Results Classification of Diseased Leaf Image

Image	Original Image	Cluster Image	Early Disease Type	Classification Result	Accuracy (%)
1			*Alternaria Alternata*	*Alternaria Alternata*	98.2
2			*Alternaria Alternata*	*Alternaria Alternata*	98.3871
3			*Alternaria Alternata*	*Cercospora Leaf Spot*	98.2
4			*Alternaria Alternata*	*Cercospora Leaf Spot*	98.3871
5			*Alternaria Alternata*	*Cercospora Leaf Spot*	98.3871
6			*Alternaria Alternata*	*Alternaria Alternata*	98.3871

7			*Anthracnose*	*Cercospora Leaf Spot*	98.3871
8			*Anthracnose*	*Anthracnose*	98.3871
9			*Anthracnose*	*Alternaria Alternata*	98.2
10			*Anthracnose*	*Cercospora Leaf Spot*	98.3871
11			*Anthracnose*	*Cercospora Leaf Spot*	98.3871
12			*Bacterial Blight*	*Cercospora Leaf Spot*	98.3871

13			*Bacterial Blight*	*Bacterial Blight*	98.2%
14			*Bacterial Blight*	*Bacterial Blight*	98.3871
15			*Bacterial Blight*	*Cercospora Leaf Spot*	98.3871
16			*Cercospora Leaf Spot*	*Cercospora Leaf Spot*	98.3871
17			*Cercospora Leaf Spot*	*Cercospora Leaf Spot*	98.3871
18			*Cercospora Leaf Spot*	*Cercospora Leaf Spot*	98.3871

19			*Cercospora Leaf Spot*	*Cercospora Leaf Spot*	98.3871
20			*Cercospora Leaf Spot*	*Cercospora Leaf Spot*	98.3871

With the above research, the classification of each type of leaf disease, such as Alternaria Alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot by using clusters detected by leaf diseases. The results of the average level of accuracy in this study amounted to 98,349%.

CLOSING

Conclusion

Detecting leaf diseases such as Alternaria Alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot can use image processing techniques. Techniques in image processing such as image acquisition, pre-processing, segmentation, feature extraction, and classification. Based on the results of experiments and analyzes on the leaf disease classification system based on texture using texture feature extraction and Support Vector Machine.

Although the results of classifying the types of leaf diseases do not all get good results, but with the accuracy of the performance classification Support Vector Machine managed to get an average high level of accuracy of 98,349%.

Suggestion

Applications made have many shortcomings that can still be developed and improved. In addition to classifying leaf disease by leaf texture using the Support Vector

Machine, it can be further developed using other classification methods such as KNN, PCA, ANN, and PNN to obtain more accurate accuracy.

REFERENCES

Coley-Smith, J. R. 2008. ***Plant pathology***. *Physiological Plant Pathology*, 1(1), p. 83. doi: 10.1016/0048-4059(71)90043-9.

Muchtar, Mutmainnah dan Cahyani Laili. 2016. ***Klasifikasi Citra Daun dengan Metode Gabor Co-Occurence***, *Jurnal ULTIMA Computing*, 7(2), pp. 39–47. doi: 10.31937/sk.v7i2.231.

Vidyashanakara, M. Naveena dan Kumar, G. Hemnatha. 2018. ***Leaf Classification Based on GLCM Texture and SVM***, *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4(3), pp. 156–159.

Sollapure, Rashmi, Karadiguddi Laxmi, Hanasi Priyanka, Daddi Soumya dan Kale Bharati. 2018. ***Plant Leaf Disease Detection Based on Image Processing Using MATLAB***, pp. 5544–5548. doi: 10.15680/IJIRCCE.2018.0605085.

Asana, I Made Dwi Putra, Widyantara I Made Oka, Wirastuti, NMAED dan Adnyana Ida Bagus Putu. 2017. ***Metode Contrast Stretching untuk Perbaikan Kualitas Citra pada Proses Segmentasi Video***, *Majalah Ilmiah Teknologi Elektro*, 16(1). doi: 10.24843/mite.2017.v16i02p01.

Min, Yin dan Htun Nay Chi. 2018. ***Plant Leaf Disese Detection and Classification using Image Processing***, *International Journal of Research and Engineering*, 5(9), 516-523. doi:10.21276/ijre.2018.5.9.4.

Haralick, Robert M., Shanmugam Karthikeyan dan Dinstein, Its'Hak. 1973. ***Textural Features for Image Classification***, *IEEE Transactions on Systems, Man and Cybernetics* SMC-3, 610–621. doi:10.1109/TSMC.1973.4309314.

Albregtsen, Fritz. 2008. ***Statistical Texture Measures Computed from Gray Level Coocurrence Matrices***, *Image Processing Laboratory*, *Department of Informatics*, *Universitas of Oslo*, doi: 10.2307/302397.

Rakhmawati, Puji Utami, Pranoto Yuliana Melita dan Setyati Endang. 2018. ***Klasifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur Dan Fitur Warna Menggunakan Support Vector Machine***, pp. 1–8.

Sahaduta, Yegar. dan Lubis Chairisni. 2013. ***Gray Level Coocurence Matrix sebagai Pengekstraksi***, pp. 33–38.

Stipaničev, Darko. 1994. ***Introduction to Digital Image Processing and Analysis***, Switzerland: Chronolab Vision, Chronolab A.G. Zug Switzerland, pp. 1–78. doi: 10.1201/b15731.

Mirjalili, Vahid dan Raschka Sebastian. 2017. ***Phyton Machine Learning: Perform Python Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow***. ISBN: 9781787125933.

Kulinavar, Pooja dan Hadimani Vidya I. 2017. ***Classification of leaf disease based on Multiclass SVM classifier***, 3, pp. 321–327. ISSN: 2454-132X